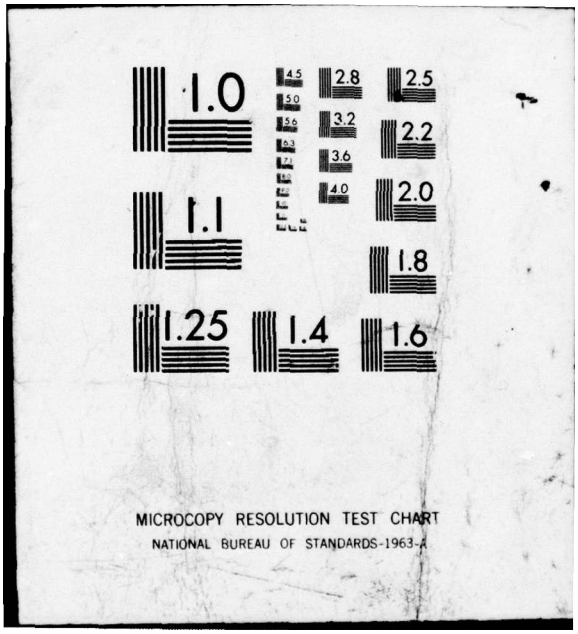
AD-A033 312 TEXAS UNIV AT AUSTIN DEPT OF ELECTRICAL ENGINEERING F/G 12/1
ANOTHER LOOK AT THE EDITED NEAREST NEIGHBOR RULE.(U)
OCT 76 C S PENROD, T J WAGNER F44620-70-C-0091
UNCLASSIFIED AFOSR-TR-76-1217 NL

| OF |
AD
A033312

END

DATE
FILMED
1-77


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ANOTHER LOOK AT THE EDITED NEAREST NEIGHBOR RULE

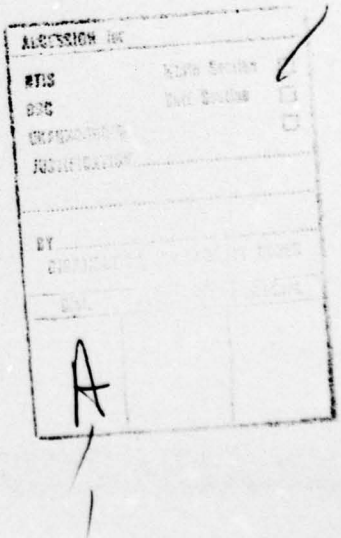
C.S. Penrod* and T.J. Wagner†
Department of Electrical Engineering
The University of Texas at Austin
Austin, Texas 78712

* Supported by JSEP Contract F44620-70-C-0091.
† Supported by AFOSR Grant 72-2371.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

Abstract

In this paper we present a slight modification of Wilson's Edited Nearest Neighbor Rule [1] in the one dimensional case for which it is possible to compute tight bounds on the average asymptotic risk. It is pointed out that the argument used by Wilson to establish his bounds is probably incorrect with the bounds being somewhat optimistic. The rule presented here is not in itself of any great significance since it does not generalize to more than one dimension. The contribution lies in the fact that for this type of rule (which is very similar to Wilson's rule) an exact analysis is possible which permits comparison of the relative merits of various editing schemes. Although no proof is offered, the strong similarities involved give reason to believe that the results concerning the relative efficiencies of the various editing schemes will carry over to higher dimensional problems with the usual version of the nearest neighbor rule.



Another Look at the Edited Nearest Neighbor Rule

Nearest neighbor rules form a widely known class of solutions to problems in the field of pattern discrimination. Several papers concerned with the various properties exhibited by these rules have appeared in the literature, most of them directing their attention toward asymptotic properties of the risk when the rule is used with a data collection of independent classified observations. Another interesting question is the following. Given data consisting of n classified observations, it will sometimes be the case that some of the observations from one class will lie in a region where most of the observations are from another class. In such a case, it may be possible to improve the rule's performance by removing from the data those observations which are "surrounded" by observations from a different class. The question is, is there an effective way to identify those observations which should be eliminated.

Wilson [1] has examined this problem and proposed the following algorithm. Take each sample of the data in turn and, using the k nearest neighbor rule with the remainder of the data, estimate its classification. The edited data set is obtained by removing from the original data set those samples which were misclassified by their k nearest neighbors. The edited nearest neighbor rule then uses the single nearest neighbor rule in conjunction with the edited data set to classify unknown observations.

Wilson used an argument to show that

$$EL_n(k) \to R^E(k)$$

where $L_n(k)$ is the conditional n sample probability of error for the edited nearest neighbor rule which uses k neighbors in the editing process, and where $R^E(k)$ satisfies

$$R^* \leq R^E(1) \leq 1.2R^*$$

$$R^* \leq R^E(3) \leq 1.149R^*$$

$$R^* \leq R^E(5) \leq 1.10R^* \ .$$

Unfortunately, the argument is incomplete. On page 413 of [1], Wilson gives an expression for $\varphi_{\infty}^{km}(1/x)$ which is claimed to be the asymptotic probability that an observation at x is assigned to class 1. Actually $\varphi_{\infty}^{km}(1/x)$ is just the proportion of samples from class 1 in a small neighborhood of x after editing. Unless those samples are uniformly distributed in the neighborhood, $\varphi_{\infty}^{km}(1/x)$ is not necessarily related to the probability that x is assigned to class 1. Wilson does not indicate any reason why the samples should be uniformly distributed and in fact intuition seems to indicate that the editing process leaves the samples distributed in clusters. (It should be noted that Tomek [2] makes the same error in arriving at his equation 13).

An exact analysis of the effect of editing on the average asymptotic performance can be done if we restrict the problem to one dimension and use a rule which selects the nearest neighbor to a point x from those samples which are greater than x. It is important to point out two things about this type of nearest neighbor rule. First, the arguments used by Cover and Hart [3] are still applicable so that these rules' asymptotic performance will be indistinguishable from that of the standard nearest neighbor rules. Second, Wilson's argument still applies to the edited version of these rules so that the same bounds arrived at in his paper would still apply, if his argument were correct. In fact, however, the analysis shows his bounds to be optimistic in this case, leaving little justification for thinking them correct in the other, more important case.

As usual, we let $(X_1, \theta_1), \ldots, (X_n, \theta_n)$ be a sequence of independent identically distributed random vectors where each observation X_j takes values in $\mathbb{R}$ and each label θ_j takes values in $\{1,2\}$. For each j,

$$P\{\theta_j = 1\} = \pi_1$$

$$P\{\theta_j = 2\} = 1 - \pi_1 = \pi_2$$

$P\{X_j \leq x \mid \theta_j = i\}$ has an almost everywhere continuous density

$$f_i,\ i = 1, 2.$$

The following two lemmas will be used in the calculation of $\varphi(1/x)$, the asymptotic probability that the nearest neighbor to x after editing is from class 1. $\varphi(1/x)$ will then be used to bound $R^E(k)$ in terms of R*.

Let x be a continuity point of f_1 and f_2, and

$$p_1(x) = P\{\theta=1 \,|\, X=x\} = \frac{\pi_1 f_1(x)}{\pi_1 f_1(x) + \pi_2 f_2(x)} ,$$

$$p_2(x) = 1 - p_1(x).$$

We will use $\theta^{(k)}$ to denote the label associated with the k^{th} nearest neighbor to x (from those samples greater than x). Finally, if $\{s_i^j\}_{i=1}^j$ is a sequence of ones and twos of length j, we will let S_j denote the event that

$$\{\theta^{(1)} = s_1^j, \ldots, \theta^{(j)} = s_j^j\}.$$

The dependence of each event S_j on n is implicit here.

Lemma. Let S_j be an event as described above, where the corresponding sequence contains m ones and j-m twos. Then

$$\lim_{n\to\infty} P(S_j) = p_1(x)^m p_2(x)^{j-m} .$$

Proof. The proof is a simple application of well known theorems concerning the convergence of the nearest neighbors to x. (See Cover and Hart [3], Wagner [4].)

Now, let $\{S_j\}_{j=1}^\infty$ be a sequence of such events, with S_j depending only on $\theta^{(1)}, \ldots, \theta^{(j)}$. We assume that S_i and S_j are disjoint for all $i \neq j$, and, if $j > n$, then S_j is empty. We also need the following easy lemma.

Lemma. Let $\{S_j\}_{j=1}^\infty$ be a sequence of events as described above. Then

$$\lim_{n\to\infty} P\Big(\bigcup_{j=1}^{\infty} S_j\Big) = \sum_{j=1}^{\infty} p_1^{\,m_j}(x)\, p_2^{\,j-m_j}(x)$$

where m_j is the number of ones in the sequence associated with S_j.

The computation of $\varphi(1/x)$ is done by specifying the sequences of labels $\theta^{(j)}$ which yield the desired result after editing. The lemmas above are then used to find the limiting probability of obtaining one of the necessary sequences.

In the case of editing with a single nearest neighbor, if the labels of the samples to the right of x are in one of the following sequences, then a class 1 sample will remain to the right of x after editing. The sequences are given as X's and O's, where an X denotes class 1, and an O denotes class 2, and where, for example, $(XO)^3$ indicates XOXOXO. The sequences of interest are $(XO)^jXX$, $j \geq 0$ and $(OX)^jX$, $j \geq 1$. By the use of the above lemmas, we can compute the limiting probability that one of the above sequences occurs as

$$\varphi_1(1/x) = \sum_{j=0}^{\infty} p_1^2 (p_1 p_2)^j + \sum_{j=1}^{\infty} p_1 (p_1 p_2)^j$$

$$= \frac{p_1^2}{1-p_1 p_2} + \frac{p_1^2 p_2}{1-p_1 p_2}$$

$$= \frac{p_1^2 (1+p_2)}{1-p_1 p_2} \quad .$$

For the case of editing with three nearest neighbors, the sequences of interest are:

XXXX
XXXO
XXOX
XOXX
$XX(OOXX)^jOX$, $j \geq 1$
$X(OOXX)^jX$, $j \geq 1$
$X(OX)^jX$, $j \geq 2$
OXXX
$O(XXOO)^jXXOX$, $j \geq 1$
$O(XO)^jXX$, $j \geq 1$
$O(XXOO)^jXXX$, $j \geq 1$
$OO(XXOO)^jXXOX$, $j \geq 1$
$OO(XXOO)^jXXX$, $j \geq 1$.

The limiting probability of obtaining one of the above sequences has been calculated to be

$$\varphi_3(1/x) = p_1^3\left[p_1+3p_2+\frac{p_2^2(1+p_1)}{1-p_1p_2}+\frac{p_2(1+p_2)[1+p_2+p_1p_2(1+p_1)]}{1-(p_1p_2)^2}\right] .$$

In the case of editing with five nearest neighbors, the analysis was done in the same fashion, but it becomes rather tedious so only the result will be given.

$$\begin{aligned}\varphi_5(1/x) = {} & p_1^6 + 6p_1^5p_2 + 15p_1^4p_2^2 + 10p_1^3p_2^3 \\ & + \frac{p_1^3p_2^3}{1-p_1p_2}\ (p_1^2 - p_2^2) \\ & + \frac{p_1^3p_2^3}{1-(p_1p_2)^3}\left[3(p_1-p_2)(p_1p_2+p_1^2p_2^2)\right. \\ & + (p_1^2-p_2^2)\ (5 + 2p_1p_2 + 4p_1^2p_2^2) + 3(p_1^3-p_2^3)(1 + p_1p_2) \\ & \left. + (p_1^4 - p_2^4)\right] .\end{aligned}$$

We also include the result obtained by editing the data set twice in succession with one nearest neighbor.

$$\varphi_{1,1}(1/x) = \frac{p_1^3}{1-2p_1p_2}\left(1 + p_2 + \frac{p_2(1-p_1^2)}{1 - p_1p_2}\right) .$$

Finally, the computation was done for editing with one nearest neighbor, then using the three nearest neighbor rule to classify unknowns. This resulted in

$$\varphi_1^3(1/x) = \frac{p_1^3}{(1-p_1p_2)^2} + \frac{p_1^2(1+p_2)}{1 - p_1p_2}\left[\frac{p_1}{1-p_1p_2} + \left(\frac{p_1p_2}{1-p_1p_2}\right)^2 - \frac{p_1^3}{(1-p_1p_2)^2}\right] .$$

These last two quantities were computed to gain some idea of the effectiveness of variations on Wilson's basic idea.

Finally, to obtain the bounds on $R^E(k)$, we compute $r_E(x)/r_B(x)$, the ratio of the local risk for the edited rule to the local risk for the Bayes rule. Note that

$$r_B(x) = \min \{p_1(x), p_2(x)\}$$

$$r_E(x) = \varphi(1/x)p_2(x) + \varphi(2/x)p_1(x)$$

$$= p_1(x) + \varphi(1/x) - 2\varphi(1/x)p_1(x) .$$

For $0 < p_1(x) \leq \frac{1}{2}$, we have

$$\frac{r_E(x)}{r_B(x)} = \frac{p_1(x) + \varphi(1/x) - 2\varphi(1/x)p_1(x)}{p_1(x)} .$$

We substitute the appropriate expressions for $\varphi(1/x)$ as computed above and find the maximum as a function of p_1. This yields

$$R^* \leq R^E(1) \leq 1.269R^*$$

$$R^* \leq R^E(3) \leq 1.204R^*$$

$$R^* \leq R^E(5) \leq 1.169R^* .$$

For comparison purposes, the bounds for the standard 1, 3, and 5 nearest neighbor rules are

$$R^* \leq R(1) \leq 2R^*$$

$$R^* \leq R(3) \leq 1.31R^*$$

$$R^* \leq R(5) \leq 1.2R^* .$$

The improvements made possible by editing the data set are obvious, although not quite as good as originally suggested by Wilson. The result obtained by editing the data twice with one nearest neighbor is

$$R^* \leq R^{2E}(1) \leq 1.162R^* .$$

Finally, editing with one neighbor followed by classifying with three neighbors yields

$$R^* \leq R_3^E(1) \leq 1.168R^* .$$

The attractiveness of the result obtained by editing twice with one nearest neighbor is diminished somewhat by the fact that it appears to be computationally more difficult to edit twice with one neighbor than to edit once with several neighbors. We note that Tomek [2] errs in stating that the opposite is true. The reason is that in either case, the distance between the samples must be computed at least once. However, in the case of editing twice, one must either store all the distances, or recompute most of them for the second edit. In the case of editing once with k neighbors, it is necessary to compute the distances only once, storing only the current k nearest neighbors as the distances are being computed.

One remaining point of interest is the amount of data reduction to be expected when the data set is edited. Asymptotically, this will depend on R_k, the asymptotic risk of the k nearest neighbor rule. We note that if we let S_n be the number of samples edited out of the data, then S_n/n is simply the deleted estimate of R_k (see Cover [5]). Wagner [6] has shown that under mild conditions satisfied here, $S_n/n \to R_k$ in probability, so asymptotically the edited data set will contain a fraction near $1 - R_k$ of the number of samples before editing. This means that in problems which have a small value of R^*, the amount of data reduction to be expected from editing is negligible.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFOSR - TR - 76 - 1217	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): ANOTHER LOOK AT THE EDITED NEAREST NEIGHBOR RULE.	5. TYPE OF REPORT & PERIOD COVERED: Interim rept.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): C.S. Penrod and T.J. Wagner	8. CONTRACT OR GRANT NUMBER(s): AFOSR 72-2371
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Texas, Department of Electrical Engineering, Austin, Texas 78712	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, D.C. 20332	12. REPORT DATE: October, 1976 13. NUMBER OF PAGES: 10
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): F44620-70-C-0091, AF-AFOSR-2371-72	15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release, distribution unlimited	11p.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)	
18. SUPPLEMENTARY NOTES	
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): pattern recognition, nonparametric discrimination	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper we present a slight modification of Wilson's Edited Nearest Neighbor Rule [1] in the one dimensional case for which it is possible to compute tight bounds on the average asymptotic risk. It is pointed out that the argument used by Wilson to establish his bounds is probably incorrect with the bounds being somewhat optimistic. The rule presented here is not in itself of any great significance since it does not generalize to more than

next page

20. Abstract (continued)

cont.

one dimension. The contribution lies in the fact that for this type of rule (which is very similar to Wilson's rule) an exact analysis is possible which permits comparison of the relative merits of various editing schemes. Although no proof is offered, the strong similarities involved give reason to believe that the results concerning the relative efficiencies of the various editing schemes will carry over to higher dimensional problems with the usual version of the nearest neighbor rule.

REFERENCES

[1] D.L. Wilson, "Asymptotic Properties of Nearest Neighbor Rules Using Edited Data," IEEE Trans. Syst. Man, Cybern., Vol. SMC-2, pp. 408-420, July 1972.

[2] I. Tomek, "A Generalization of the k-NN Rule," IEEE Trans. Syst., Man, Cybern., Vol. SMC-6, pp. 121-126, Feb. 1976.

[3] T.M. Cover and P.E. Hart, "Nearest Neighbor Pattern Classification," IEEE Trans. Inf. Th., Vol. IT-13, pp. 21-27, Jan. 1967.

[4] T.J. Wagner, "Convergence of the Nearest Neighbor Rule," IEEE Trans. Inf. Th., Vol. IT-17, pp. 566-571, Sept. 1971.

[5] T.M. Cover, "Learning in Pattern Recognition," Methodologies of Pattern Recognition, Ed. S. Watanabe, Academic Press, New York, pp. 111-132, 1969.

[6] T.J. Wagner, "Deleted Estimates of the Bayes Risk," Ann. Statist., Vol. 1, pp. 359-362, Mar. 1973.